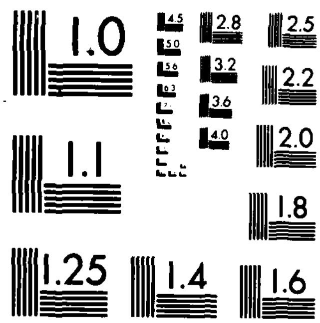
AD-A174 516 INFERENCE ON THE OCCURENCE/EXPOSURE RATE AND SIMPLE RISK RATE(U) PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS Z D BAI ET AL. AUG 86 TR-86-18 1/1

UNCLASSIFIED AFOSR-TR-86-2057 F49620-85-C-0008 F/G 6/5 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFOSR-TR- 86-2057

AD-A174 516

Approved for public release;
distribution unlimited.

Center for Multivariate Analysis

University of Pittsburgh

86 11 25 287

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 86-2057	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Inference on the occurrence/exposure rate and simple risk rate		5. TYPE OF REPORT & PERIOD COVERED Technical - August 1986
		6. PERFORMING ORG. REPORT NUMBER 86-18
7. AUTHOR(s) Z.D. Bai, P.R. Krishnaiah and Y.Q. Yin		8. CONTRACT OR GRANT NUMBER(s) F49620-85-C-0008
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Multivariate Analysis, 515 Thackeray Hall, University of Pittsburgh, Pittsburgh, PA 15260		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304 A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research, Dept. of the Air Force, Bolling Air Force Base, DC 20332 nm		12. REPORT DATE August 1986
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) same as 11.		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Asymptotic distributions, medical research, occurrence/exposure rate, reliability, risk rate

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper, ~~we~~ study$_S$ the asymptotic distributions of the functions of the occurrence/exposure rates of several groups of patients as well as Berry-Esseen bound on the distribution function of the occurrence/exposure rate. Asymptotic distributions of functions of the simple risk rates are also derived. The results are useful in not only medical research but also in the area of reliability.

DD FORM 1 JAN 73 1473

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

②

INFERENCE ON THE OCCURRENCE/EXPOSURE RATE AND SIMPLE RISK RATE*

Z.D. Bai, P.R. Krishnaiah and Y.Q. Yin

Center for Multivariate Analysis
University of Pittsburgh

August 1986

Technical Report No. 86-18

Center for Multivariate Analysis
515 Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260

* The work of the first two authors was supported by Contract F49620-85-C-0008 of the Air Force Office of Scientific Research. The United States Government is authorized to reproduce and distribute reprints for governmental purposes.

ABSTRACT

In this paper, we study the asymptotic distributions of the functions of the occurrence/exposure rates of several groups of patients as well as Berry-Esseen bound on the distribution function of the occurrence/exposure rate. Asymptotic distributions of functions of the simple risk rates are also derived. The results are useful in not only medical research but also in the area of reliability.

Key words and phrases: Asymptotic distributions, medical research, occurrence/exposure rate, reliability, risk rate.

1. INTRODUCTION

In medical studies, it is of interest to study the association between the occurrence of certain diseases and the exposure factors. Various measures of risk of a disease are considered (e.g., Breslow and Day (1980), Howe (1985)) in the literature. One such measure is the ratio of the number of patients died to the total number of individuals observed in a fixed time period. Using this measure, various authors have studied some of the statistical problems connected with the risk rate. Another measure used in the literature for the risk is the ratio of the number of persons died to the total number of years exposed to risk. For surveys of some developments on the theory of occurrence/exposure rates, the reader is referred to Berry (1983) and Hoem (1976). The main object of this paper is to study some problems connected with the occurrence/exposure measure. Some results are also obtained on risk rates.

Suppose an experiment is conducted for a fixed period of time T and n patients are observed during this period. Also, let X_i denote the total time i-th patient is exposed to risk. Then, the risk measure considered in this paper is

$$R_n = V_n/U_n \tag{1.1}$$

where $U_n = Y_1 + \ldots + Y_n$, $V_n = Z_1 + \ldots + Z_n$, and

$$Y_i = \begin{cases} X_i & \text{if } X_i \le T \\ T & \text{if } X_i > T \end{cases}$$

$$Z_i = \begin{cases} 1 & \text{if } X_i \le T \\ 0 & \text{if } X_i > T. \end{cases}$$

The denominator in (1.1) is known as person-years.

In Section 2 of this paper, we establish asymptotic normality of a function of R_n. In Section 3, we establish the Berry-Esseen bound on the distribution of R_n. This bound is quite useful since it gives an upper bound on the absolute value of the difference between the distribution functions of R_n and the normal variable with mean zero and variance one. The bound is of order $c/\sqrt{n}$ where c is a constant and n is the sample size. The asymptotic distributions of the ratios of the measures in several groups are given in Section 4. In Section 5, we consider the measure V_n/n and give results analogous to those given in Sections 3 - 4 for the measure R_n. The results of this paper are useful not only in medical research but also in the area of reliability. For example, consider the situation when n items of an equipment are under test for performance under stress over a period of time T. A measure of reliability of the equipment is the ratio of the number of items which did not fail to the total number of items under test during the period of time T. It is also of interest to find the ratio of the number of items which did not fail to $X_1 + \ldots + X_n$ where X_i denotes the duration of the time i-th item is under test.

2. ASYMPTOTIC NORMALITY OF THE OCCURRENCE/EXPOSURE RATE

Let $p = P[X_i > T] = 1-q$. If $p = 1$, then $R_n = 0$ whereas $R_n = n/(X_1 + \ldots + X_n)$ when $p = 0$. Both of the above cases are simple and so we deal the case when $p \varepsilon (0,1)$.

Using strong law of large numbers for i.i.d. sequence , we have $(V_n/n) \to q$ almost surely (a.s.) and

$$\frac{1}{n} U_n \to u = EY_1 = E(X_1)I[X_1 \leq T] + Tp \quad \text{a.s.}$$

as $n \to \infty$. Hence, $R_n \to q/u$ a.s. Now, let $W_i = uZ_i - qY_i$, $r = q/u$ and

$$\xi_n = \sqrt{n}\,(R_n - r)$$

$$= \frac{n}{uU_n}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i\right]. \tag{2.1}$$

Here $\{W_i\}$ is a sequence of bounded i.i.d. random variables with mean zero. So, by central limit theorem, we observe that $\sum_{i=1}^{n} W_i/\sqrt{n}$ is asymptotically distributed as normal with mean zero and variance σ^2, where

$$\sigma^2 = E(W_1^2) = E(uZ_1 - qY_1)^2. \tag{2.2}$$

Since $\frac{1}{n} uU_n \to u^2$ a.s., we obtain that ξ_n is asymptotically distributed as normal with mean zero and variance σ^2/u^4.

Now, let $f(\cdot)$ denote a function which is continuously differentiable for two times around r, say in $(r-\delta\ ,\ r+\delta)$, $\delta > 0$. By Taylor's expansion, if $|R_n - r| < \delta/2$, we obtain

$$\sqrt{n}\left(f(R_n)-f(r)\right)$$

$$= f'(r)\xi_n + \frac{1}{2\sqrt{n}}\ \xi_n^2 f''(\zeta_n),$$

where ζ_n is a number between r and R_n. Because f'' is bounded in the interval $(r-\delta/2\ ,\ r+\delta/2)$, ξ_n tends to a normal variable in distribution and $P(|R_n - r| \geq \delta/2) \to 0$, and we have the following theorem.

<u>THEOREM 1</u>

Under the condition mentioned above,

$$\sqrt{n}\left(f(R_n)-f(r)\right) \to N\left(0,\ \left(f'(r)\right)^2\sigma^2/u^4\right).$$

In practice, the asymptotic variance of $\sqrt{n}\left(f(R_n)-f(r)\right)$ is unknown. In such situations, we use the following approximate confidence interval on $f(r)$:

$$\left|\sqrt{n}\left(f(R_n)-f(r)\right)\right| \leq d_\alpha a(f)$$

where $a(f)$ can be taken as

$$\left(\sqrt{n}\,|f'(R_n)|/U_n^2\right) \sqrt{\sum_{j=1}^{n}\left(\sum_{i=1}^{n}(Y_i Z_j - Y_j Z_i)\right)^2}$$

which is a consistent estimate of $|f'(r)|\sigma/u^2$ and d_α is the upper $100\alpha\%$ point of the normal distribution with mean zero and variance one.

3. BERRY-ESSEEN BOUND FOR THE DISTRIBUTION OF THE OCCURRENCE/EXPOSURE RATE

Let

$$\eta_n = \frac{u^2}{\sigma} \xi_n = \frac{nu}{U_n} \cdot \frac{1}{\sqrt{n}\,\sigma} \sum_{i=1}^{n} W_i .$$

Then, according to the result proved in previous section, η_n is asymptotically distributed as normal with mean zero and variance one. Let F_n denote the distribution function of η_n and Φ that of the standard normal. In this section, we shall prove the following.

THEOREM 2

There exists a constant c such that

$$||F_n - \Phi|| = \sup_x |F_n(x) - \Phi(x)| \le c/\sqrt{n} , \tag{3.1}$$

where Φ is the standard normal distribution function. In the sequel, we need the following lemma.

LEMMA 1

Let $\{X_n, Y_n, Z_n\}$ be a sequence of random vectors with relation $X_n = Y_n + Z_n$ and let F_n, G_n denote the distribution functions of X_n and Y_n respectively. If there exist constants c_i, $i = 1,2,3$, such that

1) $||G_n - \Phi|| \le c_1/\sqrt{n}$.

2) $P(|Z_n| \ge c_2/\sqrt{n}) \le c_3/\sqrt{n}$,

then there exists a constant c_4 such that

$$||F_n - \Phi|| \le c_4/\sqrt{n} .$$

For a proof of the above lemma, the reader is referred to X.R. Chen (1981).

Now, we turn to prove Theorem 2. Let

$$\frac{nu}{U_n} = 1 + \frac{1}{n}\sum_{i=1}^{n} (1-(Y_i/u)) + \Delta_n \tag{3.2}$$

Then

$$\eta_n = S_n + \Delta_n' + \Delta_n'' + \Delta_n''' + \Delta_n'''' , \tag{3.3}$$

where

$$S_n = \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n} W_i , \tag{3.4}$$

$$\Delta_n' = \frac{1}{n^{3/2}\sigma}\sum_{1\le i\ne j\le m} W_j(1-Y_i/u), \tag{3.5}$$

$$\Delta_n'' = \frac{1}{n^{3/2}\sigma}\sum_{i=1}^{n} W_i(1-Y_i/u), \tag{3.6}$$

$$\Delta_n''' = \frac{1}{n^{3/2}\sigma}\sum\nolimits_1 W_j(1-Y_i/u), \tag{3.7}$$

$$\Delta_n'''' = \Delta_n S_n , \tag{3.8}$$

and $m = n - \sqrt{n}$, the summation $\sum_1$ runs over all possible values of i and j such that $1\le i\le n$, $m+1\le j\le n$, $i\ne j$ or $1\le j\le n$, $m+1\le i\le n$, $i\ne j$.

At first, we see that

$$P(|\Delta_n'''| \ge \frac{1}{\sqrt{n}}) \le nE(\Delta_n''')^2$$

$$= \sigma^{-2}n^{-2}\sum\nolimits_1[EW_j^2(1-Y_i/u)^2 + 2EW_iW_j(1-Y_i/u)(1-Y_j/u)]$$

$$\le 3\sigma^{-2}n^{-1/2}\; EW_1^2(1-Y_2/u)^2$$

$$\le c/\sqrt{n}, \tag{3.9}$$

where and in the sequel c denotes positive constant but may take different value at each appearance. Also, for any $c \geq \sigma^{-1}\left(|EW_1(1-Y_1/u)|+1\right)$, we have

$$P\left(|\Delta_n''| \geq c/\sqrt{n}\right)$$

$$= P\left(|\sum_{i=1}^{n} W_i(1-Y_i/u)| \geq c\sigma n\right)$$

$$\leq P\left(|\sum_{i=1}^{n} W_i(1-Y_i/u) - EW_1(1-Y_1/u)| \geq n\right)$$

$$\leq n^{-1}\mathrm{Var}\left(W_1(1-Y_1/u)\right) \leq c/\sqrt{n}\ . \tag{3.10}$$

We now estimate Δ_n''''. Define the event

$$E_n = \{|\frac{1}{n}\sum_{i=1}^{n}(1-Y_i/u)| \geq \frac{1}{2}\}$$

By Hoeffding inequality (see Hoeffding (1963)), we have

$$P(E_n) \leq 2\exp\{-2n(1/2T)^2\}. \tag{3.11}$$

Let E_n^c denote the complement of the event E_n. When E_n^c is true, we have

$$|\Delta_n| = |\sum_{k=2}^{\infty}\left(\frac{1}{n}\sum_{i=1}^{n}(1-Y_i/u)\right)^k| \leq 2\left(\frac{1}{n}\sum_{i=1}^{n}(1-Y_i/u)\right)^2.$$

Thus

$$P(|\Delta_n''''| \geq 1/\sqrt{n})$$

$$= P(|\Delta_n S_n| \geq 1/\sqrt{n})$$

$$\leq P(E_n) + P(E_n^c,\ |S_n\Delta_n| \geq 1/\sqrt{n})$$

$$\leq P(E_n) + P\left(2\left(\frac{1}{n}\sum_{i=1}^{n}(1-Y_i/u)\right)^2 \left|\frac{1}{\sqrt{n}\,\sigma}\sum_{i=1}^{n} W_i\right| \geq 1/\sqrt{n}\right)$$

$$\leq P(E_n) + P\left(\left|\frac{1}{n}\sum_{i=1}^{n}(1-Y_i/u)\right| \geq \frac{1}{\sqrt{2}}\, n^{-\frac{3}{8}}\right) + P\left(\left|\frac{1}{\sqrt{n}\,\sigma}\sum_{i=1}^{n} W_i\right| \geq n^{\frac{1}{4}}\right) \quad (3.12)$$

By Hoeffding inequality, we get

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}(1-Y_i/u)\right| \geq \frac{1}{\sqrt{2}}\, n^{-3/8}\right)$$

$$\leq 2\exp\{-2n(\frac{u}{\sqrt{2}\,T}\, n^{-3/8})^2\}$$

$$\leq c/\sqrt{n} \quad (3.13)$$

and

$$P\left(\left|\frac{1}{\sqrt{n}\,\sigma}\sum_{i=1}^{n} W_i\right| \geq n^{1/4}\right)$$

$$\leq 2\exp\{-2n\left((\frac{\sigma}{T+1})\, n^{-1/4})\right)^2\} \leq c/\sqrt{n}\ . \quad (3.14)$$

From (3.11) - (3.14), it follows that

$$P\left(|\Delta_n''''| \geq 1/\sqrt{n}\right) \leq c/\sqrt{n}$$

Applying Lemma 3.1, to prove Theorem 2, we only need to prove that

$$||G_n - \Phi|| \leq c/\sqrt{n}\ , \quad (3.15)$$

where G_n denotes the distribution function of $T_n = S_n + \Delta_n'$, and S_n, Δ_n' were defined in (3.4), (3.5).

Now, write

$$f_n(t) = E\exp\{itS_n\},$$

$$\tilde{f}_n(t) = E\exp\{itT_n\},$$

$$a_i = \frac{1}{\sigma} W_i,$$

$$b_{ij} = \frac{1}{\sigma}(1-Y_i/u)W_j,$$

$$S_{n1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{m} a_i,$$

$$S_{n2} = \frac{1}{\sqrt{n}} \sum_{i=m+1}^{n} a_i.$$

Then, we have

$$|f_n(t)-\tilde{f}_n(t)| = |Ee^{itS_n}(e^{it\Delta_n'}-1)|$$

$$\leq |t|\ |E\Delta_n' e^{itS_n}| + \frac{t^2}{2}|E(\Delta_n')^2\theta_n e^{itS_n}|, \tag{3.16}$$

where θ_n is a complex function of $t\Delta_n'$ with $|\theta_n| \leq 1$. Hence, θ_n is independent of S_{n2}. Thus

$$|E(\Delta_n')^2\theta_n e^{itS_n}| \leq E(\Delta_n')^2|Ee^{itS_{n2}}|. \tag{3.17}$$

Now let

$$V(t) = Ee^{ita_1}.$$

Then we have

$$|V(t)| \leq \exp\{-\frac{1}{2}t^2 + \frac{2}{3}|t|^3 E|a_1|^3\} \tag{3.18}$$

(The proof of (3.18) can be found in Chapter 5 of Petrov's book (1975)). Therefore there exists a constant $\delta_1 > 0$, such that for any $|t| \le \delta_1$,

$$|v(t)| \le \exp\{-\frac{1}{4}t^2\}. \tag{3.19}$$

Hence for $|t| \le \delta_1\sqrt{n}$, we have

$$\begin{aligned}|Ee^{itS_{n2}}| &\le |v(t/\sqrt{n})|^{[\sqrt{n}]} \\ &\le \exp\{-t^2/4\sqrt{n}\}\end{aligned} \tag{3.20}$$

and

$$\begin{aligned}&|E\exp\{it(S_n-(a_1+a_2)/\sqrt{n})\}| \\ &= |v(t/\sqrt{n})|^{n-2} \le \exp\{-t^2(n-2)/4n\} \\ &\le \exp\{-t^2/5\} \quad \text{for large } n.\end{aligned} \tag{3.21}$$

Note

$$\begin{aligned}E(\Delta_n')^2 &= \frac{1}{n^3\sigma^2}\, m(m-1)[Eb_{12}^2 + 2Eb_{12}b_{21}] \\ &\le c/n.\end{aligned}$$

Hence from (3.17), we get for $|t| \le \delta_1\sqrt{n}$,

$$\left|\frac{t^2}{2}E(\Delta_n')^2\theta_n e^{itS_n}\right| \le c(t^2/n)\exp\{-t^2/4\sqrt{n}\}. \tag{3.22}$$

Now write

$$g_n(t) = Eb_{12}\exp\{it(a_1+a_2)/\sqrt{n}\}.$$

Since $Eb_{12} = 0$, $Eb_{12}a_1 = Eb_{12}a_2 = 0$, we have

$$|g_n(t)| \le \frac{t^2}{2n} E(|b_{12}|)(a_1+a_2)^2$$
$$\le ct^2/n. \tag{3.23}$$

By (3.21), (3.23), we have for $|t| \le \delta_1 \sqrt{n}$,

$$|E\Delta_n' e^{itS_n}| \le \frac{m(m-1)}{n^{3/2}\sigma} |g_n(t)|\ |E \exp\{it(S_n-(a_1+a_2)/\sqrt{n})\}|$$
$$\le Ct^2 n^{-1/2} \exp\{-t^2/5\}. \tag{3.24}$$

From (3.16), (3.22) and (3.24), we get

$$|f_n(t)-\tilde{f}_n(t)| \le c(\frac{|t|^3}{\sqrt{n}} e^{-t^2/5} + \frac{t^2}{n} e^{-t^2/4\sqrt{n}}). \tag{3.25}$$

By lemma 1 in Chapter 5 of Petrov (1975), we have for $|t| \le \delta_2\sqrt{n}$, $\delta_2 > 0$,

$$|f_n(t)-e^{-t^2/2}| \le c \frac{|t|^3}{\sqrt{n}} e^{-t^2/8}. \tag{3.26}$$

Thus (3.25) and (3.26) yield for $|t| \le \delta \sqrt{n}$

$$|\tilde{f}_n(t)-e^{-t^2/2}| \le c[(\frac{|t|^3}{\sqrt{n}} e^{-t^2/8} + \frac{t^2}{n} e^{-t^2/4\sqrt{n}}] \tag{3.27}$$

where $\delta = \min(\delta_1,\delta_2) > 0$. From (3.27), it follows that

$$\int_{|t|\le\delta\sqrt{n}} \frac{1}{|t|} |\tilde{f}_n(t)-e^{-t^2/2}|dt$$

$$\le \frac{c}{\sqrt{n}} \int_{-\infty}^{\infty} t^2 \quad e^{-t^2/8}dt + \frac{c}{n} \int_{-\infty}^{\infty} |t|e^{-t^2/4\sqrt{n}} dt$$

$$\le \frac{c}{\sqrt{n}} .$$

Here the estimate of the last integral can be obtained by making variable transformation $u = tn^{-1/4}$. Then using Berry-Esseen's basic inequality, we prove (3.15). This completes the proof of Theorem 2.

4. ASYMPTOTIC JOINT DISTRIBUTION OF FUNCTIONS OF OCCURRENCE/EXPOSURE RATES

Let $X_1^{(j)},\ldots,X_{n_j}^{(j)}$, $(j=1,2,\ldots,s)$ be a sample drawn from the j-th population where $X_i^{(j)}$ denotes the observation on i-th individual in j-th population. Also, let

$$Y_i^{(j)} = \begin{cases} X_i^{(j)} & \text{if} \quad X_i^{(j)} \le T \\ T & \text{otherwise} \end{cases}$$

$$Z_i^{(j)} = \begin{cases} 1 & \text{if} \quad X_i^{(j)} \le T \\ 0 & \text{otherwise,} \end{cases}$$

for $j = 1,2,\ldots,s$ and $i=1,2,\ldots,n_j$. Now, let

$$R_{n_j}^{(j)} = V_{n_j}^{(j)}/U_{n_j}^{(j)} \tag{4.1}$$

for $j=1,2,\ldots,s$, where

$$U_{n_j}^{(j)} = \sum_{i=1}^{n_j} Y_i^{(j)} \ , \ V_{n_j}^{(j)} = \sum_{i=1}^{n_j} Z_i^{(j)}. \tag{4.2}$$

We know that

$$R_{n_j}^{(j)} \to r_j \ \text{ a.s. } \ j = 1,2,\ldots,s \ . \tag{4.3}$$

Let $f(x_1,x_2,\ldots,x_s)$ be a function which is continuously differentiable for two times in a neighborhood of $(r_1,\ldots,r_s)$. Suppose that

$$n/n_j \to \lambda_j < \infty \ , \text{ as } \ n \to \infty, \tag{4.4}$$

where $n = n_1 +,\ldots,+ n_j$.

Then

$$\sqrt{n}\left(f(R_{n_1}^{(1)}, R_{n_2}^{(2)},\ldots,R_{n_s}^{(s)}) - f(r_1, r_2,\ldots,r_s)\right)$$

$$= \sum_{j=1}^{s} a_j\sqrt{\frac{n}{n_j}}\,\xi_{n_j}^{(j)} + \sum_{j=1}^{s}\sum_{k=1}^{s}\sqrt{\frac{n}{n_j n_k}}\,a_{jk}\xi_{n_j}^{(j)}\xi_{n_k}^{(k)}. \tag{4.5}$$

when $R_{n_1}^{(1)}, R_{n_2}^{(2)},\ldots,R_{n_s}^{(s)}$ falls in the neighborhood of $(r_1,\ldots,r_s)$ in which f is differentiable. Here

$$a_j = \left.\frac{\partial f(x_1,\ldots,x_s)}{\partial x_j}\right|_{(x_1,\ldots,x_s) = (r_1,\ldots,r_s)}, \quad j = 1,2,\ldots,s,$$

$$a_{jk} = \left.\frac{\partial^2 f(x_1,\ldots,x_s)}{\partial x_j\partial x_k}\right|_{(x_1,\ldots,x_s) = (t_1,\ldots,t_s)}, \quad j,k = 1,2,\ldots,s$$

and $(t_1,\ldots,t_s)$ is some point on the linear section joining $R_{n_1}^{(1)},\ldots,R_{n_s}^{(s)}$ and $(r_1,\ldots,r_s)$. Let B be a non-trivial closed ball with center $(r_1,\ldots,r_s)$ which is contained in that neighborhood of $(r_1,\ldots,r_s)$. Then,

$$P\left((R_{n_1}^{(1)},\ldots,R_{n_s}^{(s)}) \notin B\right) \to 0.$$

Since

$$|a_{jk}| \le M$$

for all $j,k = 1,2,\ldots,s$ and some M when $(R_{n_1}^{(1)},\ldots,R_{n_s}^{(s)}) \in B$, we obtain

$$P\left(\left|\sum_{j=1}^{s}\sum_{k=1}^{s} a_{jk}\sqrt{\frac{n}{n_j n_k}}\;\xi_{n_j}^{(j)}\xi_{k}^{(k)}\right| \ge \varepsilon\right)$$

$$\le P\left((R_{n_1}^{(1)},\ldots,R_{n_s}^{(s)})\notin B\right) + \sum_{j=1}^{s}\sum_{k=1}^{s} P\left(|\xi_{n_j}^{(j)}\xi_{n_k}^{(k)}| \ge \frac{\varepsilon}{M}\sqrt{\frac{n_j n_k}{n}}\right) \to 0.$$

Hence

$$\sqrt{n}\left(f(R^{(1)}_{n_1},\ldots,R^{(s)}_{n_s}) - f(r_1,\ldots,r_s)\right) \to N(0\ ,\ \sigma_f^2) \qquad (4.6)$$

where

$$\sigma_f^2 = \sum_{j=1}^{s} a_j^2\ \lambda_j\ \sigma_j^2/u_j^4\ ,$$

$$u_j = E\ Y_1^{(j)},$$

$$q_j = P(X_1^{(j)} \le T)$$

and

$$\sigma_j^2 = E(W_1^{(j)})^2 = E(u_j Z_1^{(j)} - q_j Y_1^{(j)})^2.$$

Also, using the same approach to prove Theorem 2, we can establish the Berry-Esseen Bound for the distribution of $\frac{\sqrt{n}}{\sigma_f}\left(f(R^{(1)}_{n_1},\ldots,R^{(s)}_{n_s}) - f(r_1,\ldots,r_s)\right)$. The details are omitted here.

An important special case for f is $f(x_1,x_2) = x_1/x_2$. In this case, $f(R^{(1)}_{n_1},R^{(2)}_{n_2}) = R^{(1)}_{n_1}/R^{(2)}_{n_2}$ is called the ratio of occurrence/exposure rates. $R^{(1)}_{n_1}/R^{(2)}_{n_2}$ is denoted by $\widehat{RR}_n$, and we have

$$\sqrt{n_1+n_2}\ (\widehat{RR}_n - r_1/r_2) \to N(0,\ \sigma^2) \qquad (4.7)$$

where

$$\sigma^2 = a_1^2\lambda_1\sigma_1^2/u_1^4 + a_2^2\lambda_2\sigma_2^2/u_2^4$$

and

$$a_1 = 1/r_2\ ,\ a_2 = -\ r_1/r_2^2\ .$$

REMARK

Note that $R^{(2)}_{n_2}$ may be zero. However, $P(R^{(2)}_{n_2} = 0) = p_2^{n_2} \to 0$, as $n \to \infty$. Any way, the definition of RR_n for $R^{(2)}_{n_2} = 0$ does not affect the limiting result for the distribution of RR_n. However, for small sample problem, we have to make an explicit distribution of $RR_n = \infty$ when $R^{(2)}_{n_2} = 0$. Define $RR_n = 1$ when $R^{(1)}_{n_1} = R^{(2)}_{n_2} = 0$ and $RR_n = \infty$. Now let the commen density of $X_1^{(j)},\ldots,X_{n_j}^{(j)}$ be given by

$$g_j(x) = \begin{cases} \alpha_j \exp\{-\alpha_j x\} & \text{if} \quad x > 0 \\ 0 & \text{otherwise} \end{cases}$$

and let $p_j = P[X_1^{(j)} > T]$ for $j = 1,2$. We have

$$\left.\begin{aligned} P(\widehat{RR}_n=0) &= p_1^{n_1}(1-p_2^{n_2}) \\ P(\widehat{RR}_n=1) &= p_1^{n_1}p_2^{n_2} \\ P(\widehat{RR}_n=\infty) &= p_2^{n_2}(1-p_1^{n_1}) \end{aligned}\right\} \tag{4.8}$$

It is known (see Beyer, Keiding and Simonsen (1976)) that $R^{(j)}_{n_j}$ has an atom at the orgin with a mass $P_1^{n_j}$ and a density

$$f_j(x) = \sum_{k=1}^{n_j} \binom{n_j}{k} p_j^{n_j-k} \alpha_j^k T^{k-1} k x^{-2} \exp\{-\alpha_j(k-(n-k)Tx)^+/x\}$$

$$u_k[(k-(n_j-k)Tx)^+/x]\; I_{[0,\, k/(n_j-k)T]}(x), \tag{4.9}$$

for $j = 1,2$, where $I_{[a,b]}(x)$ is 1 or 0 according as x is in [a,b] or not. Hence, the distribution of RR_n, besides the three atoms given in (4.8), has a density which can be computed from the following

$$f(x) = \int_0^\infty f_1(xy) f_2(y) dy \tag{4.10}$$

Now, let

$$L_i = \sqrt{n}\ \{f_i(R_{n_1}^{(1)},\dots,R_{n_s}^{(s)}) - f_i(r_1,\dots,r_s)\}$$

for $i = 1,2,\dots,k$, $f_i(R_{n_1}^{(1)},\dots,R_{n_s}^{(s)})$ is a continuous twice-differentiable function of $R_{n_1}^{(1)},\dots,R_{n_s}^{(s)}$ around $r_1,\dots,r_s$. We have proved earlier the asymptotic normality of L_1. Following the same lines, it is easily seen that the asymptotic joint distribution of $L_1,\dots,L_k$ is multivariate normal. But the asymptotic covariance matrix of $L_1,\dots,L_k$ is isually known. We will now construct approximate confidence intervals on $f_i(r_1,\dots,r_s)$ when the covariance matrix $C = (c_{it})$ of $L_1,\dots,L_k$ is non-singular, where

$$c_{it} = \sum_{j=1}^{s} a_{i\cdot j}a_{t\cdot j}\lambda_j\sigma_j^2/u_j^4 ,$$

and

$$a_{i\cdot j} = \frac{\partial f_i(x_i,\dots,x_s)^2}{\partial x_j}\Bigg|(x_1,\dots,x_s) = (r_1,\dots,r_s)$$

In these situations, let $\hat{C}$ be a consistent estimate of C. Then $\underset{\sim}{L}'\hat{C}^{-1}\underset{\sim}{L}$ is approximately distributed as chi-square with s degrees of freedom for large samples where $L' = (L_1,\dots,L_k)$. Using this, we obtain the following approximate confidence intervals on linear combinations of $f_i(r_1,\dots,r_s)$, $(i = 1,2,\dots,k)$:

$$|\sqrt{n}\ \underset{\sim}{a}'\left(\underset{\sim}{f}(R_{n_1}^{(1)},\dots,R_{n_s}^{(s)}) - \underset{\sim}{f}(r_1,\dots,r_s)\right)| \le (g_\alpha \underset{\sim}{a}'\hat{C}\underset{\sim}{a})^{1/2}$$

for all nonnull vectors $\underset{\sim}{a}:k\times 1$ where

$$\underset{\sim}{f}(r_1,\dots,r_s) = \left(f_1(r_1,\dots,r_s)\ ,\dots,\ f_k(r_1,\dots,r_s)\right)'$$

and g_α is the upper $100\alpha\%$ point of the chi-square distribution with s degrees of freedom. The above confidence intervals are useful in constructing simultaneous confidence intervals on various ratios like

$$r_i/r_s \ (i=1,\ldots,k-1) \ , \ r_i/r_j (i < j = 2,\ldots,k) \ ,$$

$$r_i/r_{i+1} \ (i=1,2,\ldots,k-1).$$

We can also construct simultaneous confidence intervals on $f_i(r_1,\ldots,r_s)$ using Bonferroni's inequality.

5. INFERENCE ON SIMPLE RISK RATES

In this section, we compare the simple risk rates of different groups of patients who are observed for a fixed period of T years and each group may be subject to a different exposure factor. Here a simple risk rate of j-th population is defined as the proportion of individuals in that population who died during the period of observation. In this section, we use the same notation as in the preceding sections.

The sample estimate of simple risk rate for j-th population is $V_j^* = V_{n_j}^{(j)}$. Now, let $f_i(V_1^*,\ldots,V_s^*)$, $(i=1,2,\ldots,k)$, be a continuous twice differentiable function of $V_1^*,\ldots,V_s^*$ around $q_1,\ldots,q_s$.

Using Taylor's expansion, we obtain

$$L_i^* = \sqrt{n}\,\{f_i(V_1^*,\ldots,V_s^*) - f_i(q_1,\ldots,q_s)\}$$

$$= \sum_{j=1}^{s} a_{ij}\sqrt{n/n_j}\,B_j + \frac{1}{\sqrt{n}}\sum_{j=1}^{s}\sum_{k=1}^{s} a_{ijk}\sqrt{n/n_j n_k}\,B_j B_k \qquad (5.1)$$

where $B_j = \sqrt{n_j}\quad(V_{n_j}^{(j)}/n_j) - q_j$, $V_j^* = V_{n_j}^{(j)}$, and

$$a_{i\cdot j} = \left.\frac{\partial f_i}{\partial V_j^*}\right|_{\underset{\sim}{V}^* = \underset{\sim}{q}}, \quad a_{i\cdot jk} = \left.\frac{\partial^2 f_i}{\partial V_j^* \partial V_k^*}\right|_{\underset{\sim}{V}^* = \hat{\underset{\sim}{q}}}, \qquad (5.2)$$

$\underset{\sim}{V}^* = (V_1^*,\ldots,V_q^*)'$ and $\underset{\sim}{q} = (q_1,\ldots,q_s)'$ and $\hat{\underset{\sim}{q}}$ is some point on the linear section between $\underset{\sim}{q}$ and $\underset{\sim}{V}^*$. As $n \to \infty$, B_j is distributed as normal with mean 0 and variance $q_j p_j$. So, when $n, n_1,\ldots,n_s \to \infty$, the joint distribution of $L_1^*,\ldots,L_k^*$ is multivariate normal with mean vector $\underset{\sim}{0}$ and covariance matrix $C^* = (c_{it}^*)$ where

$$c_{it}^* = \sum_{j=1}^{n} a_{ij} a_{tj} \lambda_j q_j p_j \,. \qquad (5.3)$$

Let $\hat{C}^*$ be a consistent estimate of C^*. When C^* is non-singular and $n \to \infty$, we can use the following approximate simultaneous confidence intervals for the linear combinations of $q_1,\ldots,q_s$ by using the fact that $\underset{\sim}{V}^{*'}\hat{C}\underset{\sim}{V}^*$ is approximately distributed as chi-square with s degrees of freedom

$$|\sqrt{n}\ \underset{\sim}{a}'\left(\underset{\sim}{f}(V_1^*,\ldots,V_s^*) - \underset{\sim}{f}(q_1,\ldots,q_s)\right)| \le (h_\alpha \underset{\sim}{a}'\hat{C}^*\underset{\sim}{a})^{1/2} \qquad (5.4)$$

where $\underset{\sim}{f}(q_1,\ldots,q_s) = \left(f_1(q_1,\ldots,q_s)\ ,\ldots,\ f_k(q_1,\ldots,q_s)\right)'$ and h_α is the upper $100\alpha\%$ point of the chi-square distribution with k degrees of freedom.

Some special cases of $f_i(V_1^*,\ldots,V_s^*)$ are, V_1^*/V_s^*, V_i^*/V_{i+1}^*, etc. From the results given above, it is easily seen that $\sqrt{n_1+n_2}\left(\hat{R}_{12} - (q_1/q_2)\right)$ is distributed normally with mean zero and variance σ_o^2 where

$$\sigma_o^2 = (\lambda_1 q_2^2 q_1 p_1 + \lambda_2 q_1^2 q_2 p_2)/q_2^4\ ,\ \hat{R}_{12} = V_1^*/V_2^*$$

when n_1 and n_2 tend to infinity. Following similar lines as in Section 3, we can show that

$$||F_{n_1+n_2} - \Phi|| \le \frac{c}{\sqrt{n_1+n_2}}$$

where $F_{n_1+n_2}$ is the distribution function of $\sqrt{n_1+n_2}\ \sigma_o^{-1}\ \hat{R}_{12} - (q_1/q_2)$ and Φ is the distribution function of the standard normal distribution.

We know that $V_{n_j}^{(j)}$ follows the binomial distribution $B(n_j\ ,\ q_j)$, $j = 1,2,\ldots$ whatever the underlying distributions are. Hence, we have

$$P(R_{12} = x) = \begin{cases} (1-q_1)^{n_1}[1-(1-q_2)^{n_2}], & \text{if } x = 0, \\ \left(1-(1-q_1)^{n_1}\right)(1-q_2)^{n_2} & \text{if } x = \infty. \\ \sum_1 \binom{n_1}{k_1}\binom{n_2}{k_2}(1-q_1)^{n_1-k_1} q_1^{k_1}(1-q_2)^{n_2-k_2} q_2^{k_2}, & \text{otherwise.} \end{cases}$$

Here, the summation $\sum_1$ runs over all possible values of k_1 and k_2 such that $1 \le k \le n_1$, $1 \le k_2 \le n_2$ and $(k_1/n_1) = x(k_2/n_2)$ and the term for $k_1 = k_2 = 0$ appears only when $x = 1$.

If q_j is small related to n_j, $j = 1,2$, by the well-known Poisson limit theorem, we know that V_{n_j} is asymptotically distributed as Poisson distribution $P(\lambda_j)$, where $\lambda_j = n_j q_j$. Hence

$$P(\hat{R}_{12} = x) = \begin{cases} e^{-\lambda_1}(1-e^{-\lambda_2}) & \text{, if } x = 0, \\ (1-e^{-\lambda_1})\, e^{-\lambda_2} & \text{, if } x = \infty \\ \sum_2 \frac{\lambda_1^{k_1} \lambda_2^{k_2}}{k_1!\, k_2!} e^{-\lambda_1} e^{-\lambda_2} & \text{otherwise.} \end{cases}$$

Here the summation $\sum_2$ runs over all possible values of k_1 and k_2 such that $k_1 \geq 1$, $k_2 \geq 1$, $k_1/n_1 = k_2\, x/n_2$ and the term for $k_1 = k_2 = 0$ appears only when $x = 1$.

REFERENCES

[1] Berry, G. (1983). The analysis of the mortality by the subject-years method. Biometrics, 39, 173-184.

[2] Beyer, J.E., Keiding, N. and Simonsen, W. (1976). The exact behavior of the maximum likelihood estimator in the pure birth process and the pure death process. Scand. J. Statist., 3, 61-72.

[3] Breslow, N.E. and Day, N.E. (1980). Statistical Methods in Cancer Research, Volume 1: The Analysis of Case-Control Studies. International Agency for Research on Cancer, Lyon, France.

[4] Chen, X.R. (1981). Berry-Esseen bound for the error variance estimates in linear models. Sintia Sinica, No. 2, 129-140.

[5] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. 58, 13-30.

[6] Hoem, J.M. (1976). The statistical theory of demographic rates: A review of current developments. Scand. J. Statist., 3, 169-185.

[7] Howe, G.R. (1983). Confidence interval estimation for the ratio of simple and standardized rates in cohort studies. Biometrics, 39, 325-331.

[8] Petrov, V.V. (1975). Sums of Independent Random Variables. Springer-Verlong.

END

12-86

DTIC